INFORMATIONSGEHALT DER NIEDERSCHLAGSREICHE VON BUDAPEST

F. RÁKÓCZI und F. SZIDAROVSZKY

Lehrstuhl für Meteorologie L. Eötvös Univ., Budapest
Lehrstuhl für Num. und Maschin Mathematik L. Eötvös Univ., Budapest
Einegegangen am 8. 4. 1976

РЕЗЮМЕ

В работе приводится разработанная авторами техника прогноза месячной величины осадка. Техника прогноза была разработана на основе использования количества информации в последовательности данных об осадке в территории Будапешта, а методика была проверена на данных, охватывающих 10-летний период- Полученные результаты указывают на то, что с помощью предложенной техники могут быть составлены прогнозы, достоверность которых больше чем случайных совпадений.

Bei der langfristigen Vorhersage der meteorologischen Elemente greift man auf die Geschehnisse der Vergangenheit zurück and man untersucht, welche Information die in der Vergangenheit beobachteten Zeitreihen für die Zukunft darbieten können [1, 3]. Der Grundgedanke solcher Versuche ist, dass in den Wettergeschehnissen gewisse Erhaltungs- und Rückkehr-Tendenzen vorhanden sind, und dass gewisse Zirkulationscharakteristiken zu denselben oder zumindest ähnlichen Resultaten führen.

Bei einer solchen Untersuchung von Zeitreihen müssen wir zuallerest feststellen, ob nicht das Einsetzen eines Ereignisses von dem Eintritt der vorangehenden Ereignissen vollständig unabhängig sei. Wenn das Vorhandensein eines Zusammenhanges, welcher in einer gewissen Anzahl der Fälle nicht zu stark sein könnte, erwiesen wird, dann können wir hoffen, dass aus einer Kenntnis der Vergangenheit für die Zukunft Wahrscheinlichkeitsaussagen gemacht werden können. Für die Erarbeitung solcher Zusammenhänge wendet man die *Korrelationsanalyse* oder andere *Methoden der Unabhängigkeitsuntersuchungen* an, es gibt aber auch solche Verfahren, die den Informationsgehalt dieser Zeitreihen untersuchen [3], bei denen eigentlich festgestellt wird, inwieweit sich das gegebene System an die Geschehnisse der Vergangenheit „zurückzuerinnern“ fähig sei.

In der folgenden Studie werden die Monatssummen des Niederschlages von Budapest auf Grund der Beobachtungen der Jahre 1841–1960 untersucht [2].

Im Gange der Untersuchungen wird es erwogen, inwieweit die Niederschlagssumme eines gegebenen Monats die Niederschlagsverhältnisse der nachfolgenden Monate beeinflusst, und wenn ja, dann in welchem Masse. Dabei wird eigentlich mit zwei Wahrscheinlichkeitsschemas X und Y gearbeitet, deren vorherige Realisationen bekannt sind. Die Auftrittswahrscheinlichkeiten der Ereignisse der Schemas X und Y sind uns bekannt, so dass wir über die endlichen Schemas

$$X = \begin{Bmatrix} X_1, & X_2, & \dots, & X_n \\ p_1, & p_2, & \dots, & p_n \end{Bmatrix}, \qquad Y = \begin{Bmatrix} Y_1, & Y_2, & \dots, & Y_n \\ q_1, & q_2, & \dots, & q_n \end{Bmatrix}$$

verfügen können. Der simultane Eintritt der Ereignisse X_i und Y_k sei mit $X_i\ Y_k$ $(1 \leqq i \leqq n,\ 1 \leqq k \leqq m)$ bezeichnet. Die Wahrscheinlichkeit des Eintretens des zusammengesetzten Ereignisses sei W_{ik}. So erhalten wir aus dem zusammengesetzten Ereignis ein neues endliches Schema

$$[X \cdot Y] = \begin{Bmatrix} X_1 \cdot Y_1, & X_1 Y_2, & \dots & X_1 Y_m \\ X_n \cdot Y_1, & X_n Y_2, & \dots & X_n Y_m \end{Bmatrix},$$

und die Matrix der entsprechenden Wahrscheinlichkeiten in der Form

$$[P(XY)] = \begin{Bmatrix} W_{11}, & W_{12}, & \dots, & W_{1n} \\ W_{n1}, & W_{n2}, & \dots, & W_{nm} \end{Bmatrix}$$

geschrieben werden kann.

Sind X uns Y unabhängig eines vom anderen, dann besteht der Zusammenhang

$$W_{ik} = p_i q_k \,.$$

Wenn aber die Schemas X und Y voneinander abhängig sind, dann haben wir an Stelle der obigen Produktregel folgendes zu setzen:

$$W_{ik} = p_i q_{ik} \qquad (1 \leqq i \leqq n, \quad 1 \leqq k \leqq m),$$

wobei q_{ik} die Wahrscheinlichkeit des Auftretens von Y_k im Schema Y bedeutet unter der Bedingung, dass im Ereignisraum X das Ereignis X_i erfolgte.

Auf Grund der angegebenen Schemas können die Entropien $H(X)$, $H(Y)$, $H(XY)$, sowie $H_X(Y)$ und $H_Y(X)$ abgeschätzt werden, d. h. die statistische Entropie des Ereignisses X und Y, die Entropie des zusammengesetztes Ereignisses, sowie die bedingten Entropien. Im Besitze dieser kann die Informationsgrösse $I\,(X, Y)$ abgeleitet werden, welche ein Mass darüber darstellt, welchen Einfluss das Ereignis X auf die Entwicklung des Ereignisses Y ausübt. Es ist bekannt, dass

$$H(X, Y) = H(X) + H_X(Y)\,,$$

während das Informationsmass $I\,(X, Y)$ in der Form

$$I(X, Y) = H(Y) - H_X(Y)$$

geschrieben werden kann. Wenn man die zwei letzten Gleichungen addiert und eine Umordnung ausführt, kommt man zum Zusammenhang:

$$I(X, Y) = H(X) + H(Y) - H(X, Y),$$

welcher eine bequeme Möglichkeit für die Bestimmung der Informationsgrösse ermöglicht. Da im Unabhängigkeitsfall $H(X, Y) = H(X) + H(Y)$, die Abweichung von $I(X, Y)$ von Null stellt ein Mass der gegenseitigen Einwirkung eines der Ereignisse auf das andere dar.

Bei der Niederschlagsreihe von Budapest wurde, wie folgt, vorgegangen:

Die in jedem Monat des Jahres auftretenden Werte wurden in drei Wertgruppen: A_1, A_2, A_3 eingereiht. Im Gange der Eingliederung wurde das Intervall $A_{\min} - A_{\max}$ in 40 gleiche Teile aufgespaltet und die Warscheinlichkeit p_i der in die einzelnen Teilintervalle fallenden Werte untersucht, dann wurde die Summierung der Fälle soweit fortgesetzt, bis die Werte: $p_1 = 1/3$, $p_2 = 2/3$ und $p_3 = 1$ erreicht wurden. Die Werte der so bestimmten Konfidenzgrenzen – von Monat zu Monat – enthält die folgende Tabelle I.:

Tabelle I. Trennungspunkte der Intervalle

I	26.13 – 51.16
II	20.21 – 45.86
III	29.03 – 51.31
IV	36.78 – 63.90
V	52.64 – 86.98
VI	54.64 – 81.78
VII	38.27 – 65.74
VIII	35.44 – 62.54
IX	30.90 – 56.23
X	39.48 – 71.55
XI	41.24 – 67.62
XII.	36.43 – 57.87

Die Daten der Tabelle enthalten das mittlere Intervall, und wenn die Niederschlagswerte des betreffenden Monats sich zwischen diesen Grenzen befinden, wird der Monat als ein normaler Niederschlagsmonat betrachtet. Im Falle eines gegenüber der unteren Grenze niedrigeren Wertes haben wir es mit einem niederschlagsarmen Monat zu tun, während ein Monat mit einem Niederschlag über der oberen Grenze als niederschlagsreich angesehen wird.

Mit der Beachtung dieser Klassengrenzen haben wir Konfidenztabellen hergestellt und für jeden einzelnen Monat untersucht, welchen Einfluss ein jeder der vorangehenden Monate auf die Niederschlagsverhältnisse des in Rede stehenden Monats ausübt. Der Einflussgrad wurde durch die Informationsgrösse $I(X, Y)$ dargestellt. Die erhaltenen Werte der Informationsgrösse werden in der Tabelle II. gegeben:

Tabelle II.

Informationsgrössen I (X, Y) 10^4

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	79	206	136	320	429	229	151	496	263	175	428	10
2	294	56	247	441	219	272	184	297	186	44	224	116
3	499	466	315	35	153	178	137	555	597	272	39	337
4	262	277	97	164	257	102	596	125	60	414	199	150
5	70	73	396	103	385	672	395	187	79	1213	36	72
6	481	209	236	633	71	191	190	258	323	230	292	351
7	310	46	148	244	18	351	177	294	384	97	381	640
8	183	307	48	248	244	348	176	53	119	82	204	225
9	244	253	70	483	901	273	141	325	124	497	169	424
10	134	74	205	138	141	233	342	982	59	340	441	356
11	131	255	238	102	86	538	29	384	157	282	239	62

Aus den Werten der Tabelle kann man feststellen, dass in der Mehrzahl der Fälle der Zusammenhang in der Zeitfolge der Niederschlagswerte zwar nicht zu eng ist, es gibt aber für jeden Monat wenigstens drei solche vorangegangene Monate, die einen Einfluss auf die Gestaltung der Niederschlagsverhältnisse des betreffenden Monats ausüben. Es ist auch ohne weiteres klar, dass sich in der Niederschlagszeitreihe auch stärkere Einflüsse, als eine einfache Erhaltungstendenz offenbaren, sonst müssten die maximalen Werte in den Kolumnen der Tabelle II. durch die Informationsgrössen der dem betreffenden Monat vorangegangenen Monate representiert werden. Es ist auch zu ersehen, dass das zeitliche System der Zusammenhänge recht kompliziert ist, aber in der etwa halber Anzahl der Fälle die Gestaltung der Niederschlagsverhältnisse am stärksten durch die Werte der dem betreffenden Monat vorangehenden 3 – 5 Monate beeinflusst wird, während in mehreren Fällen darf man auch die Verhältnisse der vorangehenden 9 – 10 Monate nicht ausser Acht lassen. In der ersten Hälfte des Jahres – bis zum Mai – verschiebt sich der die meisten Informationen bietende Monat nach rückwerts in der Zeit, während in den Monaten Juni und Juli eine Voranrückung, erfolgt, dann haben wir mit einem unregelmässigen Verhalten zu tun. Als der von der Vorgeschichte am besten beeinflusste Monat scheint Oktober zu sein, während März als am wenigstem beeinflusst erscheint.

In der Tabelle III. – nach einer Absonderung der Monate mit den grössten Informationswerten aus der Tabelle II. – wird zusammengefasst dass für betreffenden Monat die Niederschlagsverhältnisse welchen Monats im Laufe der Schätzung der Auftrittswarscheinlichkeit in Betracht gezogen werden. Die mit Stern bezeichneten Werte in der Tabelle representieren den entsprechenden Monat des vorangehenden Jahres.

Tabelle III.

Januar	(Okt.,*	Jul.,*	Jun.*)
Februar	(Nov.,*	Okt.,*	Jun.*)
März	(Okt.,*	Dez.,*	Jan.*)
April	(Okt.,*	Jul.,*	Febr.)
Mai	(Aug.,*	Apr.,	Dez.*)
Juni	(Jan.,	Jul.,*	Nov.*)
Juli	(März,	Apr.,	Nov.*)
August	(Okt.,*	Mai.,	Jul.)
September	(Jan.,	Febr.,	März.)
Oktober	(Mai.,	Jan.,	Jun.)
November	(Jan.,	Okt.,	Apr.)
Dezember	(Mai.,	März.,	Febr.)

Aus dieser Tabeller scheint es, als ob die Niederschlagsverhältnisse des Monats Oktober den grössten Einfluss auf die anderen Monate des Jahres ausübten, i. e. diese als die besten Prediktoren benutzt werden könnten. Innerhalb des Jahres scheinen die Verhältnisse des Januars die am meisten determinierenden Werte aufzuweisen. Für das Vorzeichen des Zusammenhanges gibt das Informationsmass – seiner Definition entsprechend – keine Auskunft, so dass für die Klärung der Frage noch andere Indikatoren in Betracht genommen werden müssen. In unserem Falle werden wirt im Laufe des vorzuführenden Schätzungsverfahrens die sich auf das Vorzeichen des Zusammenhanges beziehenden Kenntisse nicht berücksichtigen, da wir nur daran interressiert sind, ob der betreffende Monat für den in Rede stehenden Monat informativ sei oder nicht. Aus den Daten der Tabelle ist es zu ersehen, dass zwischen den monatlichen Niederschlagsdaten ein verwickelter, keine einfache Markov-Kette bildender Zusammenhang anzunehmen sei. Es sei bemerkt, dass die erwähnten Zusammenhänge zwar nicht zu stark sind, sie scheinen aber genügend ausgesprochen dafür zu sein, um mit ihrer Hilfe im Besitze der Vorkenntnisse Aussagen für das Eintreten des Freignisses treffen zu können. Das hier vorzuführende klimatologische Vorhersage – Verfahren benutzt als Prediktor die in der Tabelle III. figurierenden Werte. Im Laufe der Ausarbeitung des Verfahrens wurden für jeden Monat die Grössen A_k und B_k $(1 \leqq k \leqq 12)$ bestimmt, die die Daten des betreffenden Monats in drei Gruppen teilen so, dass die Intervalle $X < A_k$, $A_k \leqq X \leqq B_k$, $X > B_k$ eine gleiche Anzahl von Beobachtungsdaten aufweisen. Die entsprechenden A_k, B_k – Werte sind in der Tabelle I. enthalten.
Im Laufe des Vorhersageverfahrens strebten wir nicht nach einer numerischen Abschätzung der monatlichen Niederschlagssumme, sondern wir wollten die Wahrscheinlichkeit anzugeben dafür dass die Monatssumme des Niederschlages kleiner als A_k ausfällt, sich zwischen A_k und B_k befindet, oder aber grösser als B_k sei.
Das Verfahren kann wie folgt zusammengefasst werden. Der Index der in die Vorhersage einbezogenen Monate sei mit $i_1, i_2, \ldots, i_r$ bezeichnet, k soll den Monat angeben, für welchen die Vorhersage ausführt werden soll.

1. Wir untersuchen es für $i = i_1, i_2, \ldots, i_r$, dass die aktuelle Angabe des i-ten Monats in welches der Intervalle $(-\infty, A_i)$, $[A_i, B_i)$, $[B_i, \infty)$ fällt.

2. Dann werden aus der beobachteten Datenreihe die Werte von Jahren aufgesucht, bei denen die Niederschlagssummen der $i_1, i_2, \ldots, i_r$-ten Monats einer nach dem anderen in dasselbe Intervall zu liegen kommen, wie im Falle des Vorhersagemonats.

3. Auf Grund dieser Jahre errechnen wir die Anzahl jener, wo die Niederschlagssummen der k-ten Monats in die Intervalle $(-\infty, A_k)$, $[A_k, B_k)$, $[B_k, \infty)$ fallen und approximieren die gesuchten Wahrscheinlichkeiten durch die relative Häufigkeiten.

Bei der Anwendung dieses Verfahrens müssen wir bei der Wahl des Parameters r eine doppelte Tendenz in Betracht nehmen. Wenn r gross ist, dann nehmen wir die Ausführung der Vorhersage auf Grund von mehreren Informationen vor. Da im Falle von r Prediktoren gibt es 3^r Fälle für die Verteilung der Daten an die drei Teilintervalle, die Anzahl der im zweiten Schritt angetroffenen Daten zu gering ausfallen und die Abschätzung der gesuchten Wahrscheinlichkeiten mit Hilfe der relativen Häufigkeiten zu ungenau sein wird. Wenn aber der Wert von r relativ klein gewählt wird, dann verlieren wir etwas an Information, demgegenüber aber werden wir im zweiten Schritt im allgemeinen mehr entsprechende Datenreihen antreffen, so dass die Wahrscheinlichkeiten im Gegensatz zum vorigen Fall durch die relativen Häufigkeiten besser approximiert werden können. Bei der Wahl des optimalen r – wertes sollen diese zwei Gesichtspunkte beachtet werden.

Die Monatssummen des Niederschlages von Budapest wurden von 1841 bis 1960 benutzt. Unser Vorhersageverfahren wurde für den Fall der nächstfolgenden zehn Jahre angewendet, um eine Kontrolle zu erhalten, und die Vorhersagen zu verifizieren. Beim Verfahren probierten wir alle drei Werte $r = 1$, $2 =$ und $r = 3$ aus. Weiterhin fürten wir die Vorhersage auch *so* aus, dass wir – die für jeden Monat berechneten Kontingenz-Tabellen anwendend – die Eintritt- wahrscheinlichkeiten der einzelnen Ereignisse in eine Rangordnung einreihten. In diesem Falle wurde als höchstwahrscheinlicher Wert diejenige Kategorie genommen, die im Laufe der 11 Vorhersagen am häufigsten den ersten Platz belegte; wenn aber so keine Entscheidung getroffen werden konnte, so beachteten wir auch den Umstand, mit welcher Häufigkeiten die betreffende Kategorie den zweiten Platz in der Rangordnung eingenommen hat.

Die Auswertung der Vorhersagen ist in der Tabelle IV. dargestellt.

Tabelle IV.

	Übereinstimmung	Fehlschlag von einer Kategorie	Fehlschlag mit zwei Kategorien
$r = 1$	42	47	31
$r = 2$	43	50	27
$r = 3$	53	43	24
	49	51	20

In der ersten Kolonne der Tabelle finden wir die Anzahl der Fälle, wo die vorhergesagten und eintreffenden Kategorien übereinstimmen. In der zweiten Kolonne wurde die Anzahl der Fälle gegeben, in welchen die vorhergesagte Kategorie in den neben der tatsächlich eingetroffenen Kategorie liegenden Ereignissraum führte. Die dritte Kolonne enthält die Anzahl der erfolglosen Vorhersagen. In der letzten Reihe der Tabelle wurden die Resultate der auf Grund der Wahrscheinlichkeitsrangordnung getroffenen Vorhersagen angeführt.

Ohne jedwede Vorhersage und ohne Benutzung der Zeitreihen sind die Wahrscheinlichkeiten eines Resultats vom ersten, zweiten, bezüglicherweise dritten Typ, eine nach der anderen: 0,333, 0,444, 0,222, da in die erste Gruppe die Ereignisse 1 – 1, 2 – 2, 3 – 3, in die zweite die Ereignisse 1 – 2, 2 – 3, 2 – 1 und 3 – 2, und in die dritte die Ereignisse 1 – 3 und 3 – 1 eingereiht wurden. Bei den von uns untersuchten 120 Fällen im Laufe der Vorhersage wächst die Anzahl der Übereinstimmungen mit dem Anwachsen der Informationsmenge, während die Anzahl der fehlgeschlagenen Prognosen abnimmt. Da bei vielen praktischen Entscheidungen als eine nützliche Information gilt die Angaben dessen, dass im gegebenen Monat eine durchschnittliche (mittlere Kategorie), eine sich unter dem Durchschnitt befindende oder aber eine überdurchschnittliche Niederschlagsmenge zu erwarten sei, so sollte die Anwendung in der langfristigen Vorhersage neben anderen, manchmal viel komplizierteren Verfahren erwogen und die Methode in die tägliche Praxis eingeführt werden.

LITERATURA

1. Nyberg, A. (1975): An experiment in forecasting monthly mean temperature in Stockholm. Tellus 27, No. 1. pp. 34 – 37.

2. Kakas, J. (Editor): Klimaatlas von Ungarn II. Publishing House of the Hung. Acad. of Sciences, Budapest.

3. Rákóczi, F. – Szidarovszky, F. (1976): Die Bedeutung von Zeitreihen bei der langfristigen Wettervorhersage Időjárás 80. 5, pp. 292 – 294